Year: 2015

Optimal structural and spectral features for tree species classification using combined airborne laser scanning and hyperspectral data

Torabzadeh, Hossein ; Leiterer, Reik ; Schaepman, Michael E ; Morsdorf, Felix

Abstract: In this paper, the performance of feature selection in tree species classification based on multi source earth observation data was studied. We applied a sequential forward floating feature selection on imaging spectroscopy (IS) and airborne laser scanning (ALS) data, as well as their combination. Qualitative comparison of the fused results shows that the selected spectral features are more distributed across the spectrum, in contrast to an accumulation of features in the near infrared region when using IS alone. A support vector machine (SVM) classifier was used for quantitative comparison of the different datasets. Assessing the classification accuracies confirmed the superiority of the selected subset of spectral and structural features compared to using all available features (improvement of > 7% in kappa accuracy).

DOI: https://doi.org/10.1109/igarss.2015.7327056

Originally published at:
Torabzadeh, Hossein; Leiterer, Reik; Schaepman, Michael E; Morsdorf, Felix (2015). Optimal structural and spectral features for tree species classification using combined airborne laser scanning and hyperspectral data. In: Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, Milan, Italy, 26 July 2015 - 31 July 2015. IEEE, 5399-5402.
DOI: https://doi.org/10.1109/igarss.2015.7327056

OPTIMAL STRUCTURAL AND SPECTRAL FEATURES FOR TREE SPECIES CLASSIFICATION USING COMBINED AIRBORNE LASER SCANNING AND HYPERSPECTRAL DATA

H. Torabzadeh, R. Leiterer, M.E. Schaepman, F. Morsdorf

Remote Sensing Laboratories, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
Email: hossein.torabzadeh, felix.morsdorf, reik.leiterer, michael.schaepman@geo.uzh.ch

ABSTRACT

In this paper, the performance of feature selection in tree species classification based on multi source earth observation data was studied. We applied a sequential forward floating feature selection on imaging spectroscopy (IS) and airborne laser scanning (ALS) data, as well as their combination. Qualitative comparison of the fused results shows that the selected spectral features are more distributed across the spectrum, in contrast to an accumulation of features in the near infrared region when using IS alone. A support vector machine (SVM) classifier was used for quantitative comparison of the different datasets. Assessing the classification accuracies confirmed the superiority of the selected subset of spectral and structural features compared to using all available features (improvement of > 7% in kappa accuracy).

Index Terms— tree species classification, sequential feature selection, imaging spectroscopy, airborne laser scanning (ALS)

1. INTRODUCTION

Spatial distribution of tree species strongly influences the quality of the forest habitats and biodiversity [1], as well as forest management [2]. Due to the difficulties and deficiencies of in-situ observations, remote sensing approaches became the first choice approach for mapping tree species in different forest ecosystems [3]. Among all remote sensing techniques, imaging spectroscopy (IS), also known as hyperspectral imaging, was used to produce quite reliable maps of species distribution [4, 5]. Nevertheless the performance of IS decreases in heterogeneous canopies (e.g. tropical and dense temperate forests), because of the high influence of the canopy structure on the measured signal [6, 7].

Fusion of IS data with airborne laser scanning (ALS) data, which provides a direct link to the structural characteristics of the forest canopy, is a contemporary approach to tackle this issue [7-10]. However, adding more features to the high-dimensional IS data aggravates the Hughes's phenomenon (or curse of dimensionality) [11] and increases the computational costs. Since the correlation between some of the structural and spectral features is high [12], a proper feature selection approach can choose the final features so that the overlap between the information content from ALS and IS data is reduced as much as possible [8]. This process ensures that we benefit most of the complementarities of ALS and IS data, but it also holds the potential to improve our understanding of the links between these data and the bio-physical/-chemical characteristics of different species.

The main objective of this study is to find the best set of ALS- and IS-derived features for tree species classification in a heterogeneous forest. For this purpose, we evaluated the results of the tree species classification using different datasets containing various subsets of the spectral/ structural features as well as all the extractable features.

2. STUDY AREA AND DATA

In this study we combine IS and ALS data in a dense temperate forest in the Swiss Jura (47°28'N, 8°21'E), which contains European beech (*Fagus sylvatica*), European ash (*Fraxinus excelsior*), Silver fir (*Abies alba*), Norway spruce (*Picea abies*), Sycamore maple (*Acer pseudoplatanus*), Wych elm (*Ulmus glabra*) and Norway maple (*Acer platanoides*), in order of dominance.

An exhaustive single tree inventory campaign was also performed in order to provide essential information (tree species, tree diameter, tree location and social status) for calibration and validation of the method and final results. IS data were acquired by the Airborne Prism Experiment (APEX) spectrometer [13] in 285 bands (400 - 2500 nm). A small footprint, full-waveform light detection and ranging (LiDAR) system was used to survey the study area during both leaf-on and leaf-off conditions, resulting in an overall echo density of about 40 points/m2. In addition, UAVdetected individual tree crown boundaries [14] were used in an object-based approach to achieve more accurate species identification rather than a simple pixel-based approach.

3. METHODS

3.1. Feature extraction

Following the pre-processing steps [12], spectral reflectance values of the sunlit part of the crowns were calculated at different wavelengths (241 spectral bands).

In addition to the ALS-derived canopy height metrics, we retrieved structural features based on the histograms of the echo height distribution (percentage of echoes per vertical bin) and the full-waveform features (e.g. summarized intensity per vertical bin) within each crown [15]. 39 structural features were derived from the leaf-on ALS data. The same amount of features was extracted from the leaf-off data, resulting in 78 structural features in total. Table 1 shows details of the ALS-derived structural features.

3.2. Feature selection

Among all the various feature selection algorithms for remotely sensed data [16-18], we used an extended sequential forward feature selection (SFS), called sequential forward floating selection (SFFS). Involving a backward procedure in SFFS allows reconsidering the previously selected features, decreasing the sensitivity to the initial conditions [19]. Together with the SFFS searching strategy, we used transformed divergence (TD) [20] as a separability evaluation criterion. The rank of each feature is indicated by the percentage of its emerging in all SFFS iterations (1276 times). The top-ranked features were then chosen so that their quantity is equal to the intrinsic dimensionality, provided by a principal component analysis (PCA). The SFFS was first applied on the IS and the ALS data separately and then on the combined IS and ALS data. The selected structural and/or spectral features subsequently formed three new datasets.

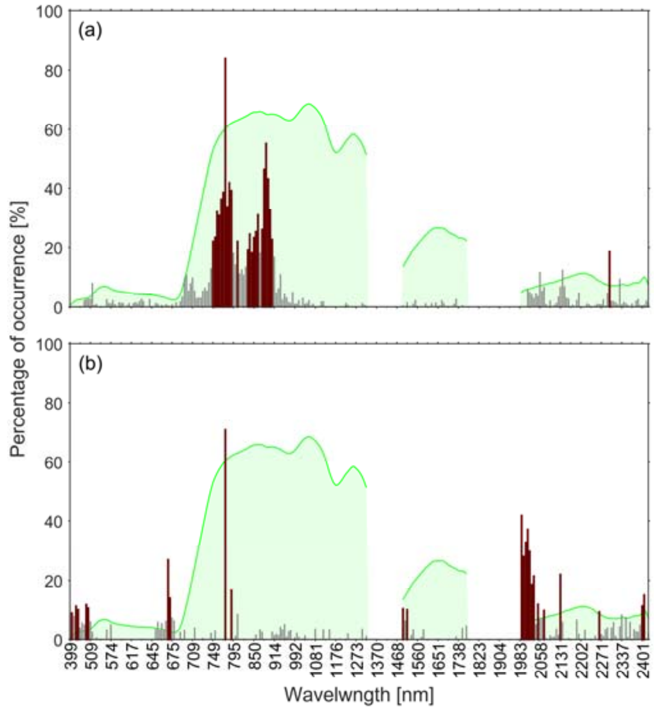

Figure 1. Selected spectral features (bands) from the IS data, (a) in absence of ALS features and (b) in presence of ALS features.

3.3. Species classification

Considering the proven advantages of support vector machine (SVM) classifiers for classifying high dimensional multiple source data [8, 21], a multiclass SVM classifier with a radial basis function kernel was deployed for distinguishing the eight most prevalent tree species in our

Table 1. ALS-derived structural features derived from ALS data in leaf-on condition. 39 similar features were calculated for leaf-off data and the conjugate differences between leaf-on and -off data were used, resulting in total 78 ALS-derived features.

Feature No.	Description
1	Maximum height of vegetation
2	Occupied length of the vertical column by vegetation
3	Ratio of vegetation length to the maximum vegetation height
4	Ratio of the height of the lowest canopy stratum to the maximum vegetation height
5	Number of detected canopy layers
6	Relative position of the largest canopy layer
7	Cumulative intensity > 1 m above ground
8	Cumulative intensity > 3m above ground
9	Cumulative intensity for the top part of the canopy
10-19	10^{th} to 90^{th} and 99^{th} height percentiles
20-29	10^{th} to 100^{th} intensity percentiles
30-39	Point density in each vertical bins (10 in total)

study area. The classifier was optimized using a five-fold cross validation during the training step. The Kappa accuracy was used for the quantitative comparison of the different classification results.

3. RESULTS AND DISCUSSION

The PCA analysis showed that the intrinsic dimensionalities of the IS and ALS data are 24 and 58, respectively. Thus, top-ranked spectral bands and structural features were separately selected (Figure 1.a). Contrastingly, using both IS and ALS datasets together resulted a different arrangement of features (Figure 1.b).

Comparison of these results reveals that in the absence of ALS-derived features the selected spectral bands are accumulated in the NIR plateau, whereas involving the structural features into the SFFS causes them to mostly disappear from this spectral region. In other words, adding the structural information resulted in a more uniform distribution of the selected spectral features across the spectrum. However, the differences between the selected ALS features are not so clear to interpret (Figure 2).

Three more datasets were created based on the feature selection results, providing six datasets for tree species classification (Table 2).

Table 2. Datasets used for classification, including different combination of spectral and structural features, along with their respective SVM classification accuracies.

No.	Spectral features	Structural features	Kappa %	Description
1	241		75.38	All the APEX bands
2		78	75.81	All the ALS-derived features
3	24		75.34	Selected spectral features
4		58	73.32	Selected structural features
5	241	78	83.71	Combination of all the spectral and the structural features
6	24	58	90.33	Combination of the selected features

The classification results showed that the achieved accuracy using the selected spectral bands (24) is very similar to the accuracy obtained by using all the spectral bands (241), even though only ten percent of the data is used. Regarding the structural features, this difference is less than 3%. Nevertheless, the achieved accuracy by the combination of the selected features is considerably higher (+7.62%) than the one resulting from all the spectral and structural features. This finding confirms that the feature selection provides more appropriate datasets for tree species classification rather than a simple combination of all the possible features.

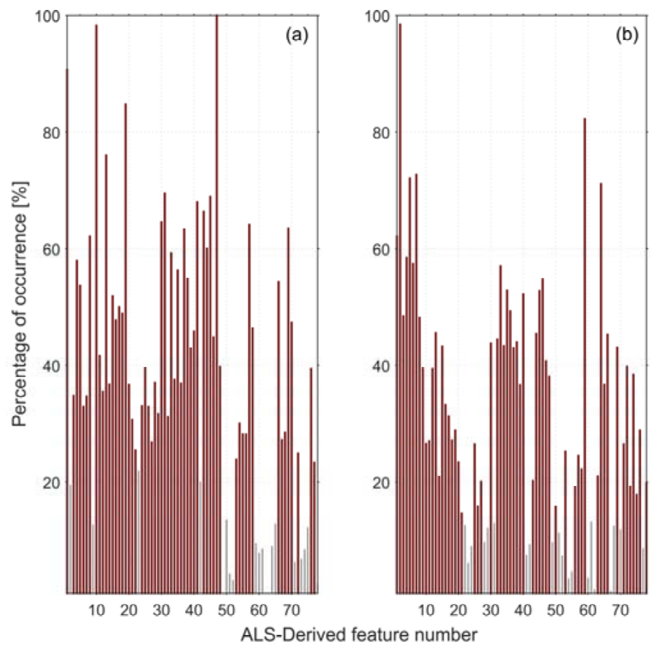

Figure 2. Selected structural features from ALS data, (a) in absence of IS features and (b) in presence of IS features.

4. CONCLUSIONS

In this study we applied a feature selection approach to IS and ALS data to find the best subset of relevant spectral and structural features for tree species classification in dense forests. Even though our results show that the selected spectral and structural features improve classification accuracy only by a few percent, the absolute increase in accuracy from about 83% to 90% could be the crucial bit making an operational application feasible. In addition, feature selection accelerates the classification process due to using less data.

5. REFERENCES

[1] J. Puumalainen, P. Kennedy, and S. Folving, "Monitoring forest biodiversity: a European perspective with reference to temperate and boreal forest zone," *Journal of Environmental Management,* vol. 67, pp. 5-14, 1/1/ 2003.

[2] H. O. Ørka, M. Dalponte, T. Gobakken, E. Næsset, and L. T. Ene, "Characterizing forest species composition using multiple remote sensing data sources and inventory approaches," *Scandinavian Journal of Forest Research,* vol. 28, pp. 677-688, 2013/10/01 2013.

[3] M. A. Wulder, J. C. White, S. N. Goward, J. G. Masek, J. R. Irons, M. Herold*, et al.*, "Landsat continuity: Issues and

opportunities for land cover monitoring," *Remote Sensing of Environment,* vol. 112, pp. 955-969, 3/18/ 2008.

[4] M. A. Cochrane, "Using vegetation reflectance variability for species level classification of hyperspectral data," *International Journal of Remote Sensing,* vol. 21, pp. 2075-2087, 2000/01/01 2000.

[5] J. Feret and G. P. Asner, "Tree Species Discrimination in Tropical Forests Using Airborne Imaging Spectroscopy," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 51, pp. 73-84, 2013.

[6] D. G. Leckie, S. Tinis, T. Nelson, C. Burnett, F. A. Gougeon, E. Cloney*, et al.*, "Issues in species classification of trees in old growth conifer stands," *Canadian Journal of Remote Sensing,* vol. 31, pp. 175-190, 2005/01/01 2005.

[7] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sensing of Environment,* vol. 96, pp. 375-398, 6/30/ 2005.

[8] M. Dalponte, L. Bruzzone, and D. Gianelle, "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data," *Remote Sensing of Environment,* vol. 123, pp. 258-270, 2012.

[9] H. Torabzadeh, F. Morsdorf, and M. E. Schaepman, "Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems – A review," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 97, pp. 25-35, 11// 2014.

[10] B. Koetz, F. Morsdorf, S. Vanderlinden, T. Curt, and B. Allgower, "Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data," *Forest Ecology and Management,* vol. 256, pp. 263-271, 2008.

[11] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *Information Theory, IEEE Transactions on,* vol. 14, pp. 55-63, 1968.

[12] H. Torabzadeh, R. Leiterer, F. Morsdorf, and M. E. Schaepman, "Tree species discrimination in dense temperate mixed forests using combined imaging spectroscopic and airborne laser scanning data," *Remote Sensing of Environment,* 2015 (in revision).

[13] M. E. Schaepman, M. Jehle, A. Hueni, P. D'Odorico, A. Damm, J. Weyermann*, et al.*, "Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX)," *Remote Sensing of Environment,* vol. 158, pp. 207-219, 3/1/ 2015.

[14] H. Torabzadeh, F. Morsdorf, R. Leiterer, and M. E. Schaepman, "Fusing imaging spectrometry and airborne laser scanning data for tree species discrimination," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, 2014, pp. 1253-1256.

[15] R. Leiterer, W. Mücke, F. Morsdorf, M. Hollaus, N. Pfeifer, and M. E. Schaepman, "Operational forest structure monitoring using airborne laser scanning," *Photogrammetrie, Fernerkundung, Geoinformation,* vol. 2013, pp. 173-184, 2013.

[16] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognition,* vol. 33, pp. 25-41, 1// 2000.

[17] M. Pal and G. M. Foody, "Feature Selection for Classification of Hyperspectral Data by SVM," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 48, pp. 2297-2307, 2010.

[18] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 39, pp. 1360-1367, 2001.

[19] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters,* vol. 15, pp. 1119-1125, 11// 1994.

[20] G. Forestier, J. Inglada, C. Wemmert, and P. Gançarski, "Comparison of optical sensors discrimination ability using spectral libraries," *International Journal of Remote Sensing,* vol. 34, pp. 2327-2349, 2013/04/10 2012.

[21] T. G. Jones, N. C. Coops, and T. Sharma, "Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada," *Remote Sensing of Environment,* vol. 114, pp. 2841-2852, 2010.